

## LOGICAL STRUCTURE SENTENCE RETRIEVAL SYSTEM

**Publication number:** JP7044579 (A)

**Publication date:** 1995-02-14

**Inventor(s):** IWASAKI MASAJIROU

**Applicant(s):** RICOH KK

**Classification:**

- international: **G06F12/00; G06F12/14; G06F17/21; G06F17/27; G06F17/30; G06F21/24; G06F12/00; G06F12/14; G06F17/21; G06F17/27; G06F17/30; G06F21/00; (IPC1-7): G06F17/30; G06F12/00; G06F17/27**

- European:

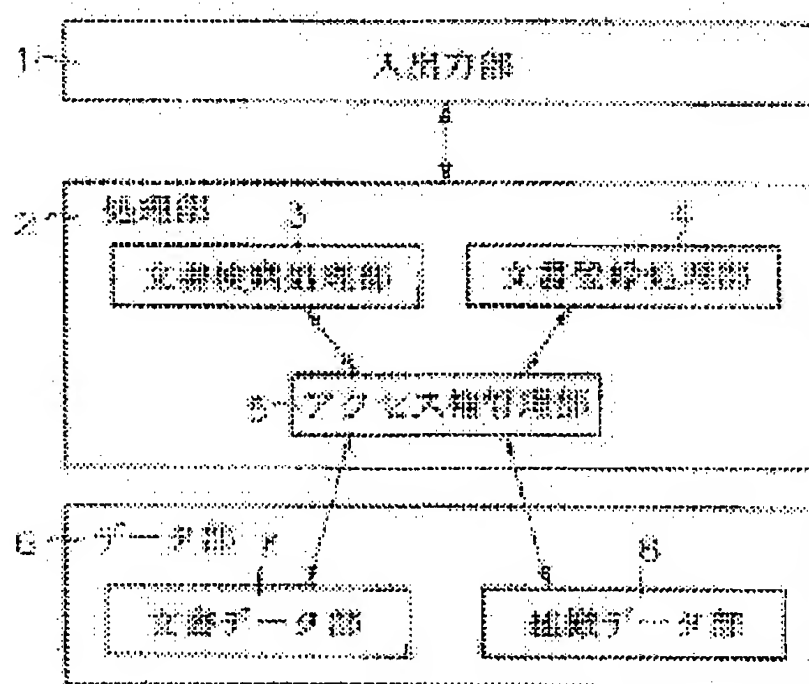
**Application number:** JP19940110690 19940525

**Priority number(s):** JP19940110690 19940525; JP19930146909 19930525

### Abstract of JP 7044579 (A)

**PURPOSE:** To register a document having a logical structure by automatically resolving it into each logical structure, and to attain the precise setting of an access right by combining texture data.

**CONSTITUTION:** A document registering request inputted to an inputting and outputting part 1 is processed by a document registration processing part 4 of a processing part 2, and registered through an access right managing part 5 in a document data part 7 of a data part 6. Also, a retrieving request inputted to the inputting and outputting part 1 is processed by a document retrieval processing part 3, the access right is checked by the access right managing part 5 based on the data of a texture data part 8, and the data are retrieved from the document data part 7, and outputted to the inputting and outputting part 1. Thus, at the time of registering the document having the logical structure, the document is automatically divided into the logical structure units, and registered in a data base with the structure information, and the retrieval depending on the structure information can be attained by each logical structure unit.



Data supplied from the **esp@cenet** database — Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-44579

(43) 公開日 平成7年(1995)2月14日

(51) Int.Cl. <sup>8</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30 12/00 17/27	5 3 7 A	8944-5B		
		9194-5L 7315-5L	G 0 6 F 15/ 419 15/ 20	3 1 0 5 5 0 E
審査請求 未請求 請求項の数15 O L (全 10 頁) 最終頁に続く				

(21) 出願番号 特願平6-110690

(22) 出願日 平成6年(1994)5月25日

(31) 優先権主張番号 特願平5-146909

(32) 優先日 平5(1993)5月25日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 岩崎 雅二郎

東京都大田区中馬込1丁目3番6号 株式  
会社リコー内

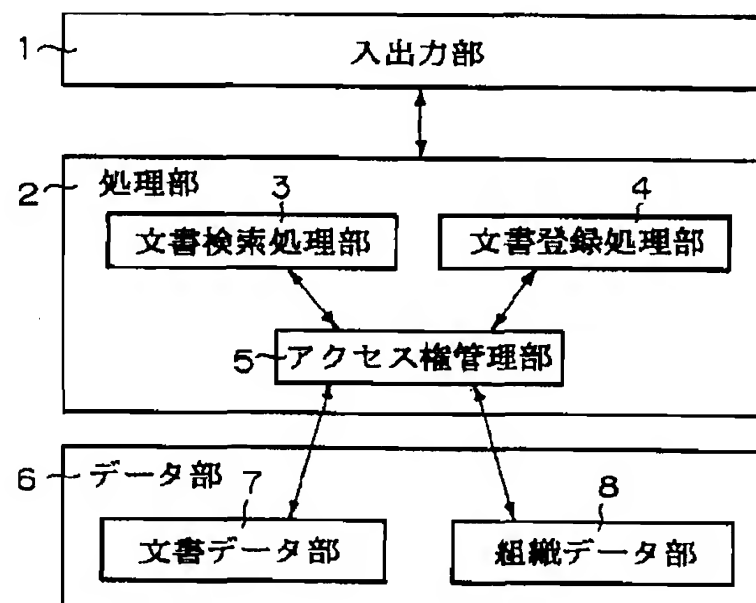
(74) 代理人 弁理士 高野 明近 (外1名)

(54) 【発明の名称】 論理構造文書検索方式

(57) 【要約】

【目的】 論理構造をもった文書を、自動的に各論理構造に分解して登録ができ、組織データを組み合わせることのできる細かなアクセス権の設定を可能とする。

【構成】 入出力部1に入力された文書登録要求は、処理部2の文書登録処理4で処理され、アクセス権管理5を介してデータ部6の文書データ7に登録される。また、入出力部1で入力された検索要求は、文書検索処理部3で処理され、組織データ8のデータを基にアクセス権管理5でアクセス権をチェックし、文書データ7からデータを検索し、入出力部1に出力する。このような構成により、論理構造を有する文書を登録する時に、自動的に論理構造単位に分割し、構造情報と共にデータベースに登録し、個々の論理構造単位に構造情報に依存する検索ができる。



## 【特許請求の範囲】

【請求項 1】 論理構造を有する文書を保管検索するデータベースにおいて、論理構造を有する文書を登録する時に、自動的に論理構造単位に分割し、構造情報と共にデータベースに登録し、個々の論理構造単位に構造情報に依存する検索ができるようにしたことを特徴とする論理構造文書検索方式。

【請求項 2】 前記文書の論理構造である木構造での位置情報を各論理構造に付与することにより、該論理構造の木構造に依存する検索を処理できることを特徴とする請求項 1 記載の論理構造文書検索方式。

【請求項 3】 前記論理構造単位にアクセス権情報を設定してアクセス権管理を行なうことを特徴とする請求項 1 記載の論理構造文書検索方式。

【請求項 4】 組織データを管理するデータベースと連動し、該組織データを利用してアクセス権の対象ユーザを設定でき、該対象ユーザを自由に変更できるようにし、前記組織データの構成が変更されてもエレメント単位のアクセス権情報形式を変更することなく対応することを特徴とする請求項 3 記載の論理構造文書検索方式。

【請求項 5】 前記アクセス権の対象ユーザが木構造を有する場合に、該木構造上のあるユーザより上位に位置するユーザすべてをアクセス権の対象とするような木構造に依存するユーザ指定ができることを特徴とする請求項 4 記載の論理構造文書検索方式。

【請求項 6】 前記アクセス権の変更をする権限を分離することによりセキュリティを高めることを特徴とする請求項 3 記載の論理構造文書検索方式。

【請求項 7】 前記各論理構造に付与されているアクセス権が実際に効力のある実効アクセス権として使われるのではなく、該実効アクセスの決定ルールが論理構造単位に設定でき、該決定ルールによって実効アクセス権を決定することを特徴とする請求項 3 記載の論理構造文書検索方式。

【請求項 8】 前記各論理構造に設定されたアクセス権及び木構造における他の論理構造のアクセス権との関係により実効アクセス権が決定されることを特徴とする請求項 1 記載の論理構造文書検索方式。

【請求項 9】 論理構造を有する文書やその文書から利用されるイメージデータのような文書以外のデータも保管検索できる構造化文書データベースにおいて、文書論理構造のタイプごとにスキーマを作成するのではなく、データベース側では文書論理構造のタイプを意識せず文書論理構造のタイプの変更及び拡張に対して柔軟に対応できるように、論理構造の最小単位である論理構造単位を共通のスキーマとして利用することを特徴とする論理構造文書検索方式。

【請求項 10】 構造化文書の文字データのみを抽出し、文字列検索データベースに抽出した文字データを登録し、検索時に文字列検索データベースを利用すること

により論理構造単位の文字列検索を可能とすることを特徴とする請求項 9 記載の論理構造文書検索方式。

【請求項 11】 ある特定の論理構造に包含されるすべての論理構造のデータを対象とした検索を高速化するために、予めシステムに登録した論理構造名が文書中に現れた場合には、該論理構造及び包含されるすべての論理構造から文字データを抽出して文字列検索データベースに登録し、特定の論理構造のデータを検索する時には前記文字列検索データベースを利用することを特徴とする請求項 9 記載の論理構造文書検索方式。

【請求項 12】 イメージデータなどの検索において、該イメージデータを利用している文書中の図題及び図の意味を表す文字データを検索することにより、前記イメージデータを適切に検索することを特徴とする請求項 9 記載の論理構造文書検索方式。

【請求項 13】 文書中でそのイメージデータに付与されている図題などの文字データを文書登録時に抽出し、イメージデータの検索用のインデックスとして登録することにより、イメージデータを高速に検索することを特徴とする請求項 9 又は 10 記載の論理構造文書検索方式。

【請求項 14】 イメージデータが複数の文書から参照されている場合には、参照するすべての文書中の図題を抽出し、イメージデータの検索用インデックスに付与することにより、イメージデータの検索洩れを減らすことを特徴とする請求項 9 又は 10 記載の論理構造文書検索方式。

【請求項 15】 文書が複数の文書から参照されている場合には、参照数が多いほど重要度が高いと想定し、文書検索時に検索結果とともに重要度を検索者に提示することで適切な検索を可能とすることを特徴とする請求項 9 記載の論理構造文書検索方式。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、論理構造文書検索方式に関し、より詳細には、文書が各章から構成され、各章が段落から構成されるというような論理的な構造を有する文書を管理する文書検索システムにおける検索及びアクセス権に係る論理構造文書検索方式に関する。

## 【0002】

【従来の技術】文書の記憶及び管理を行う文書データベース管理システムにおけるアクセス権管理方式は、例えば、特開平 2 - 2 6 6 4 4 5 号公報に提案されている。この公報のものは、データベース管理手段が文書データ格納手段へのアクセス権を得るための条件を記憶したアクセス権管理手段を有し、該アクセス権管理手段に記憶された条件を満たした場合は、前記文書データ格納手段へのアクセスを許可するようにしたものである。すなわち、あらかじめ文書の論理構造の各構造単位にデータが分割されているデータを記憶し、各構造単位にアクセス

権を設定する方式である。

#### 【0003】

【発明が解決しようとする課題】このような従来の方法では、文書単位の水、または既に文書の構造単位に分割されているデータを個別にデータベースに記憶することは可能であるが、文書自体が文書構造情報を含む文書を自動的に論理構造に分割して、登録することができない。また、登録した文書についての論理構造に依存する検索を十分ではなく、アクセス権の対象ユーザ及びパーミッションについても細かな設定ができないという問題点があった。さらに、登録した文書についての論理構造に依存する検索や文書中で利用されているイメージデータの検索機能が不十分であるという問題点があった。

【0004】本発明は、上述のごとき実情に鑑みなされたもので、論理構造をもった文書を自動的に各論理構造に分解して登録ができ、組織データと組み合わせることにより、きめ細かなアクセス権の設定が可能であるようにすること、また、論理構造をもった文書を自動的に各論理構造に分解し登録ができ、新たに文書タイプを追加する場合でもスキーマ自体を変更する必要がなく、また、論理構造に関連する検索やイメージの検索を高速にかつ適切に処理することができるようにした論理構造文書検索方式を提供することを目的としてなされたものである。

#### 【0005】

【課題を解決するための手段】本発明は、上記目的を達成するために、(1) 論理構造(文書は各章から構成され各章は段落から構成されるというような論理的な構造)を有する文書を保管検索するデータベースにおいて、論理構造を有する文書を登録する時に、自動的に論理構造単位に分割し、構造情報と共にデータベースに登録し、個々の論理構造単位に構造情報に依存する検索ができるようにすること、更には、(2) 前記文書の論理構造である木構造での位置情報を各論理構造に付与することにより、該論理構造の木構造に依存する検索を処理できること、更には、(3) 前記論理構造単位にアクセス権情報を設定してアクセス権管理を行なうこと、更には、(4) 前記(3)において、組織データ(ユーザ、組織、グループなどの利用者の様々な情報)を管理するデータベースと連動し、該組織データを利用してアクセス権の対象ユーザを設定でき、該対象ユーザを自由に変更できるようにし、前記組織データの構成が変更されてもエレメント単位のアクセス権情報形式を変更することなく対応すること、更には、(5) 前記(4)において、前記アクセス権の対象ユーザが木構造を有する場合に、該木構造上のあるユーザより上位に位置するユーザすべてをアクセス権の対象とするような木構造に依存するユーザ指定ができること、更には、(6) 前記(3)において、前記アクセス権の変更をする権限を分離することによりセキュリティを高めること、更には、(7)

前記(3)において、前記各論理構造に付与されているアクセス権が実際に効力のある実効アクセス権として使われるのではなく、該実効アクセスの決定ルールが論理構造単位に設定でき、該決定ルールによって実効アクセス権を決定すること、更には、(8) 前記(1)において、前記各論理構造に設定されたアクセス権及び木構造における他の論理構造のアクセス権との関係により実効アクセス権が決定されること、(9) 論理構造(文書は各章から構成され各章は段落から構成されるというような論理的な構造)を有する文書(構造化文書)やその文書から利用されるイメージデータのような文書以外のデータも保管検索できる構造化文書データベースにおいて、文書論理構造のタイプ(文書タイプ)ごとにスキーマを作成するのではなく、データベース側では文書論理構造のタイプを意識せず文書論理構造のタイプの変更及び拡張に対して柔軟に対応できるように、論理構造の最小単位である論理構造単位を共通のスキーマとして利用すること、更には、(10) 前記(9)において、構造化文書の文字データのみを抽出し、文字列検索データベースに抽出した文字データを登録し、検索時に文字列検索データベースを利用することにより論理構造単位の文字列検索を可能とすること、更には、(11) 前記

(9)において、ある特定の論理構造に包含されるすべての論理構造のデータを対象とした検索を高速化するために、予めシステムに登録した論理構造名が文書中に現れた場合には、該論理構造及び包含されるすべての論理構造から文字データを抽出して文字列検索データベースに登録し、特定の論理構造のデータを検索する時には前記文字列検索データベースを利用すること、更には、

(12) 前記(9)において、イメージデータなどの検索において、該イメージデータを利用している文書中の図題及び図の意味を表す文字データを検索することにより、前記イメージデータを適切に検索すること、更には、(13) 前記(9)又は(10)において、文書中でそのイメージデータに付与されている図題などの文字データを文書登録時に抽出し、イメージデータの検索用のインデックスとして登録することにより、イメージデータを高速に検索すること、更には、(14) 前記

(9)又は(10)において、イメージデータが複数の文書から参照されている場合には、参照するすべての文書中の図題を抽出し、イメージデータの検索用インデックスに付与することにより、イメージデータの検索洩れを減らすこと、更には、(15) 前記(9)において、文書(イメージデータなどを含む)が複数の文書から参照されている場合には、参照数が多いほど重要度(参照数)が高いと想定し、文書(イメージ)検索時に検索結果とともに重要度を検索者に提示することで適切な検索を可能とすることを特徴としたものである。

#### 【0006】

【作用】入出力部に入力された文書登録要求は、処理部

の文書登録処理で処理され、アクセス権管理を介してデータ部の文書データに登録される。また、入出力部で入力された検索要求は、文書検索処理部で処理され、組織データのデータを基にアクセス権管理でアクセス権をチェックし、文書データからデータを検索し、入出力部に出力する。このような構成により、論理構造を有する文書を登録する時に、自動的に論理構造単位に分割し、構造情報と共にデータベースに登録し、個々の論理構造単位に構造情報に依存する検索ができる。

【0007】また、入出力部に入力された文書登録要求は、処理部の文書登録処理で処理され、アクセス権管理を介しデータ部の構造化文書データに登録される。また、入出力部で入力された検索要求は、文書検索処理部で処理され、組織データのデータを基にアクセス権管理でアクセス権をチェックし、文書データからデータを検索し、入出力部に出力する。文書登録処理部及び文書検索処理部は文字列検索データベースを論理構造検索のために利用する。このような構成により、論理構造をもった文書を自動的に各論理構造に分解し登録ができ、新たに文書タイプを追加する場合でもスキーマ自体を変更する必要がなく、また、論理構造に関連する検索やイメージの検索を高速にかつ適切に処理することができる。

#### 【0008】

【実施例】実施例について、図面を参照して以下に説明する。図1は、本発明による論理構造文書検索方式の一実施例を説明するための構成図で、図中、1は入出力部、2は処理部、3は文書検索処理部、4は文書登録処理部、5はアクセス権管理部、6はデータ部、7は文書データ部、8は組織データ部である。

【0009】入出力部1に入力された文書登録要求は、処理部2の文書登録処理4で処理され、アクセス権管理5を介してデータ部6の文書データ7に登録される。また、入出力部1で入力された検索要求は、文書検索処理部3で処理され、組織データ8のデータを基にアクセス権管理5でアクセス権をチェックし、文書データ7からデータを検索し、入出力部1に出力する。

【0010】以下に、本発明による論理構造文書検索方式の各機能について順次説明する。

#### 登録

SGMLなどの文書記述言語では、図2で示すように、会議開催通知は件名、作成者、記事から構成され、さらに、記事は日時、場所、主催から構成されるというような文書の論理構造を記述できる。このような論理構造をもつ文書を登録する時に、自動的に論理構造単位（エレメント）に分解し、それぞれをデータベースに登録する。登録時に各エレメントに付与される情報には次のようなものがある。

- 【0011】・エレメントの名前：件名、作成者など
- ・エレメントの内容：件名の内容、作成者の名前など
- ・エレメントのID

- ・パーミッション（アクセス権情報）
- ・親エレメントの名前
- ・親エレメントID
- ・エレメントパス

したがって、これらの情報により、エレメント単位に検索が可能である。

#### 【0012】検索

前述のエレメントごとに付与されている情報により、例えば、作成者がXXXさんの文書を検索したい場合には、エレメントの名前が“作成者”で、エレメントの内容が“XXX”であるエレメントを検索することができる。また、各エレメントは木構造を構成することができるように親エレメントのIDをもつので、親エレメントのIDを辿ることによって木構造に依存する検索が可能であるが、検索速度は遅い。そこで本発明は木構造に依存する検索を高速に処理できるように、各エレメントはエレメントパスを有する。エレメントパスとは木構造のルートのエレメントから辿ったパスである。エレメントパスは、UNIXのディレクトリ指定と同様に“/”で各エレメントが区切られている。例えば、図2の“場所”のエレメントパスは“/会議開催通知/記事/場所”となる。このエレメントパスを利用して高速に検索ができる。

#### 【0013】アクセス権情報

各エレメント（論理構造単位）はアクセス権情報（パーミッション）を有し、エレメント単位にアクセス権の管理が可能である。アクセス権情報は次のような情報からなり、各エレメントに複数のアクセス権情報を設定できる。

- ・パーミッション対象ユーザ指定

ーオブジェクトID：組織データのオブジェクトを示すID。

ーメンバID：オブジェクト内での個々のメンバを示すID。

ー対象ユーザ集合決定ルール：パーミッション対象ユーザの集合を決定するルール

- 【0014】・パーミッション：エレメントに与えられるパーミッション。次のようなパーミッションがある。

ー読出し

ー更新

ーパーミッション読出し

ーパーミッションの更新

ー挿入

ー削除

実際に有効なパーミッションは、次の実効アクセス権決定ルールに従い決定する。

- ・実効アクセス権決定ルール：実効アクセス権決定ルールに従い、実際に有効なアクセス権が決定される。

それぞれについて以下に詳細に述べる。

- 50 【0015】パーミッション対象ユーザ指定



組織データは、図 3 のような構成を有する。組織データで管理されるデータのオブジェクト ID をユーザ = 1、組織 = 2、グループ = 3、部屋 = 4 とする。また、メンバー ID をユーザ B = 2、A 事業部 = 10 とすると、ユーザ B はオブジェクト ID = 1、メンバー ID = 2 で示される。また、A 事業部はオブジェクト ID = 2、メンバー ID = 10 で示される。

【0016】さらに、対象ユーザ集合決定ルールには次のようなものがあり、これに従って、最終的な対象ユーザが決められる。

①オブジェクト ID 及びメンバー ID で指定されたメンバーに含まれるユーザを対象ユーザ集合とする。

②オブジェクト ID 及びメンバー ID で指定されたメンバーの上位のすべてのメンバーに含まれるユーザをすべて対象ユーザ集合とする。

③オブジェクト ID 及びメンバー ID で指定されたメンバーの下位のすべてのメンバーに含まれるユーザをすべて対象ユーザ集合とする。

【0017】この対象ユーザ集合決定ルールの中からいずれかを選択する。例えば、前述のオブジェクト ID 及びメンバー ID で指定された組織（メンバー）が B 事業部で、メンバー決定ルールが③の場合には、A 事業部の下の組織である B 部及び C 部も含まれる。したがって、ユーザ A、ユーザ B、ユーザ C はアクセス権対象ユーザとなる（図 3）。

【0018】従来のアクセス権管理では対象ユーザが固定である。例えば、UNIX のファイルシステムではユーザとグループのパーミッションの指定フィールドが固定になっており、パーミッションの指定に組織を加えようとすることは不可能である。これに対し、対象ユーザの指定が前述のパーミッション対象ユーザ指定で自由に設定できる。また、組織データを利用してアクセス権を管理することができるので、極めて細かいアクセス権の設定が可能である。

#### 【0019】パーミッション

パーミッションの読み出し、書き込みの権利を文書データの読み出し書き込みの権利と独立し、セキュリティの向上を図っている。したがって、データの書き込みはできるが、パーミッションは変更できないという設定も可能である。

#### 【0020】実効アクセス権決定ルール

前述のように、各エレメントにはアクセス権が設定されているが、実際のアクセス権（実効アクセス権）は木構造における他のエレメントに付与されているアクセス権及び各エレメントに付与されている実効アクセス権決定ルールによって生成される。実行アクセス権決定ルールには次のようなものがある。

①他のエレメントのアクセス権の影響を受けない。

②親エレメントのアクセス権が優先される。

③子エレメントのアクセス権が優先される。

④ルートエレメントのアクセス権が優先される。

実効アクセス権決定ルールが、前記②や③のような場合には、エレメントに設定されたアクセス権とは異なり、木構造での他のエレメントに設定されたアクセス権が実効アクセス権となることがある。

【0021】このように、エレメント単位に実行アクセス権決定ルールが設定できることにより、例えば、文書全体のアクセス権を書き込み不可にした場合に、あるエレメントは、文書全体のアクセス権の変更に伴い書き込み不可になるが、あるエレメントは文書全体のアクセス権に関わらず書き込み可にするという指定が可能である。

【0022】各エレメントには複数のアクセス権情報が付与でき、パーミッションを決定するために、各アクセス権情報について次のような処理をする。

①対象ユーザ決定ルールを基にオブジェクト ID 及びメンバー ID により組織データベースのデータを検索し、ユーザが含まれているかどうかを調べる。含まれていないならば、次のアクセス権情報を調べ、ユーザが含まれるアクセス権情報が見つかるまで繰り返す。すべてのアクセス権情報を調べて、ユーザが該当するアクセス権情報がないならパーミッションがないとする。

②エレメントに付与されているパーミッション及びパーミッション決定ルールに伴いパーミッションを決定する。

【0023】図 4 は、本発明による論理構造文書検索方式の他の実施例を説明するための構成図で、図中、11 は入出力部、12 は処理部、13 は文書検索処理部、14 は文書登録処理部、15 はアクセス権管理部、16 はデータ部、17 は構造化文書データ、18 は組織データ、19 は文字列検索データベース、20 は非構造化文書データ、21 は論理構造ポインタテーブルである。

【0024】入出力部 11 に入力された文書登録要求は、処理部 12 の文書登録処理部 14 で処理され、アクセス権管理部 15 を介してデータ部 16 の構造化文書データ 17 に登録される。また、入出力部 11 で入力された検索要求は文書検索処理部 13 で処理され、組織データ 18 のデータ 11 を基にアクセス権管理部 15 でアクセス権をチェックし、文書データ 17 からデータを検索し、入出力部 11 に出力する。文書登録処理部 14 及び文書検索処理部 13 は文字列検索データベース 19 を論理構造検索のために利用する。

【0025】SGML などの文書記述言語では、図 5 に示すように、会議開催通知は件名、作成者、記事から構成され、さらに、記事は日時、場所、主催から構成されるというような文書の論理構造を記述できる。このような論理構造をもつ文書を登録する時に、自動的に論理構造単位に分解してそれぞれをデータベースに登録する。各論理構造単位は、図のように包含関係及び順序関係を示すために相互にポインタをもち、このポインタによっ

て論理構造を構成することができる。また、各論理構造単位にアクセス権管理のためのパーミッション情報をもつ。

【0026】文書タイプごとにスキーマを生成するのではなく、論理構造単位に共通のスキーマ構造を定義することにより、文書のタイプが異なる場合にも新たにスキーマを定義する必要がなく柔軟な構造化文書データベースを可能とする。また、論理構造の内容に関する検索を高速にするために、文書登録時に論理構造単位のデータを文字列検索データベースに登録し、検索時には文字列検索データベースによって高速な検索を実現する。

【0027】しかし、実際には、論理構造単位は包含関係を持ち、論理構造名の本来の文字データは包含されている下位の論理構造単位に含まれている場合が多い。したがって、前述の検索だけでは高速な検索ができない。例えば、論理構造に依存した検索の例として「章」の論理構造単位に文字列検索を行ないたい場合がある。つまり「構造化文書」という言葉について検索する時に、どの章に出現するかを知りたい場合がある。このような場合にはすべての論理構造単位のデータをサーチして「構造化文書」というキーワードがあるか否かを調べなければならない。さらに「章」の論理構造は複雑な論理構造単位から成り立っているため、処理が極めて複雑になり、検索速度が遅い。

【0028】そこで、検索対象とする論理構造単位を予めシステムに登録し、文書登録時にその論理構造単位及び包含される論理構造単位のすべての文字データを文字列検索データベースにも登録することにより、包含される論理構造単位も含め高速に検索することを可能とする。

【0029】構造化文書データベースでは構造化文書だけでなく、文書中で使われるイメージデータを保管検索する。これは非構造化文書として、図6に示すように、構造化文書データベースに蓄えられる。イメージデータのみを対象としたデータベースの場合には、通常イメージデータにキーワードを予め付与し、そのキーワードを基に検索する。しかし、イメージデータにキーワードを付与する作業を必要とするだけでなく、適切なキーワードが付与されていなければ結局適切な検索ができない。

【0030】構造化文書中で使われるイメージデータは、文書中で通常図題などの形でタイトルが付けられており、イメージデータの内容を適切に表現している。しかも、構造化文書の場合、そのタイトルのデータを抽出することが容易である。そこで、キーワードを付与する代わりにタイトルの文字列データをキーワードとして利用する。

#### 【0031】登録

構造化文書は文書構造の解析を行ない、各論理構造単位に、図7に示すように文書データベースの論理構造内容検索データに一文書として登録する。登録したデータベ

ース上の文書IDと論理構造へのポインタのペアを論理構造ポインタに登録する。そして、文字列検索システムにおける文書IDを設定する。さらに、予めシステムに設定された論理構造単位的高速検索を可能とするために、システムに登録されている論理構造名ならばその論理構造単位に含まれるすべての論理構造単位の文字列をすべて抽出し、文字列検索システムの特定期理構造内容検索データに一文書として登録する。登録したデータベース上の文書IDと論理構造へのポインタのペアを論理構造ポインタテーブルに登録する。そして論理構造データに文字列検索システムにおける文書IDを設定する。

【0032】また、文書中に使われる図などのイメージデータは、予め構造化文書データベースに登録しておき、登録時に参照しているイメージデータが存在するならばそのイメージデータに付与されている図題を抽出する。図6に示すように、抽出したデータをイメージデータのタイトルデータテーブルに登録する。イメージデータが複数の文書から参照される場合には、複数のタイトルデータが登録できるようにテーブル形式となっている。複数のデータを登録することによって適切な検索を可能とする。なお、タイトルデータのデータ量が多い場合には、文字列検索システムに登録することにより、検索を高速化することができる。

#### 【0033】検索

##### (1) 論理構造名検索

論理構造単位の検索を行なう場合には、論理構造単位名と論理構造単位データを検索することによって実現できる。例えば、作成者が×××さんの文書を検索したい場合には、論理構造単位の名前が“作成者”で、論理構造単位の内容が“×××”である論理構造単位を検索すれば良い。

##### 【0034】(2) 構造依存検索

また、文書内での構造検索は、前述の包含関係及び順序関係を示すポインタを辿ることにより、様々な構造依存の検索が可能である。例えば、章単位に作成者が異なり、第1章の作成者を検索したい場合には、前述の論理構造名検索ではすべての作成者が検索されてしまう。そこで、ルートの論理構造単位からポインタを辿り1番目の章の論理構造単位を求め、そこから作成者を辿り検索する。

##### 【0035】(3) 論理構造内容検索

論理構造単位の内容検索としては、例えば、「構造化文書」という文字列を検索する場合には、文字列検索データベースの論理構造内容検索データにより、「構造化文書」という文字列を検索して文書IDを得る。この文書IDにより論理構造ポインタテーブルを引き、結果として「構造化文書」という文字列を含む論理構造が得られる。

##### 【0036】(4) 特定期理構造内容検索

予めシステムに登録されている論理構造については、特

定論理構造内容検索が可能である。この場合、前述の論理構造内容検索とは異なり、包含されるすべての論理構造の文字データが対象となる。論理構造内容検索と処理は同じである。ただし、文字列検索データベースの特定論理構造内容検索データを用いる点異なる。

#### 【0037】 (5) イメージデータ検索

イメージデータの検索は任意の文字列を検索語とする。タイトルデータをサーチし、入力された検索語を含むイメージデータを検索結果とする。また、検索結果を表示する時にイメージデータに付与されているタイトルデータの個数(参照する文書数)を重要度とし、重要度の高い順に検索者に提示する。こうすることによって、検索者は重要度の高いデータを効率良く検索できる。

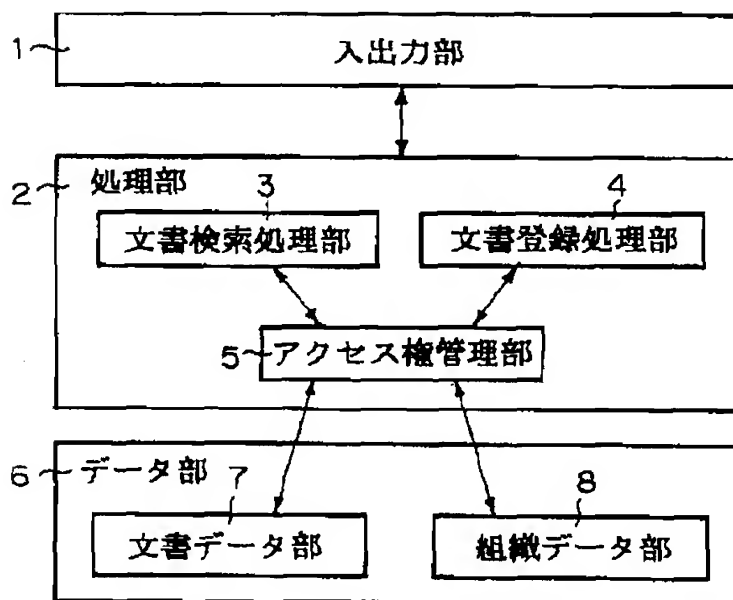
#### 【0038】

【発明の効果】以上の説明から明らかなように、本発明によると、以下のような効果がある。

(1) 論理構造(文書は各章から構成され各章は段落から構成されるというような論理的な構造)を有する文書を保管検索するデータベースにおいて、論理構造を有する文書を登録する時に、論理構造をもった文書を自動的に各論理構造に分解して登録ができ、組織データと組み合わせることによりきめ細かなアクセス権の設定が可能となる。

(2) 論理構造をもった文書を自動的に各論理構造に分解して登録ができ、新たに文書タイプを追加する場合でもスキーマ自体を変更する必要がない。また、論理構造

【図1】



に関連する検索やイメージの検索を高速にかつ適切に処理することができる。

#### 【図面の簡単な説明】

【図1】 本発明による論理構造文書検索方式の一実施例を説明するための構成図である。

【図2】 本発明におけるSGML文書の例を示す図である。

【図3】 本発明における組織データの例を示す図である。

10 【図4】 本発明による論理構造文書検索方式の他の実施例を説明するための構成図である。

【図5】 本発明におけるSGML文書の他の例を示す図である。

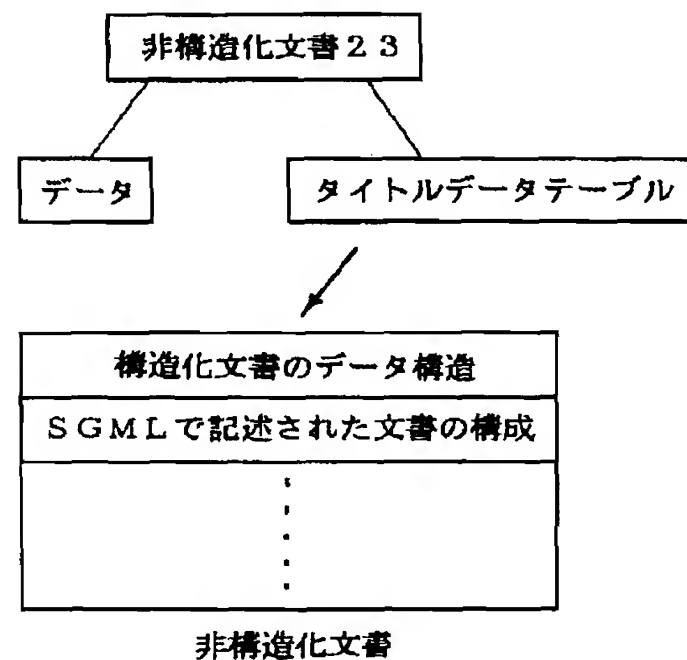
【図6】 本発明における非構造化文書を示す図である。

【図7】 本発明における論理構造単位検索のためのデータ構造を示す図である。

#### 【符号の説明】

1…入出力部、2…処理部、3…文書検索処理部、4…文書登録処理部、5…アクセス権管理部、6…データ部、7…文書データ部、8…組織データ部、11…入出力部、12…処理部、13…文書検索処理部、14…文書登録処理部、15…アクセス権管理部、16…データ部、17…構造化文書データ、18…組織データ、19…文字列検索データベース、20…非構造化文書データ、21…論理構造ポインタテーブル。

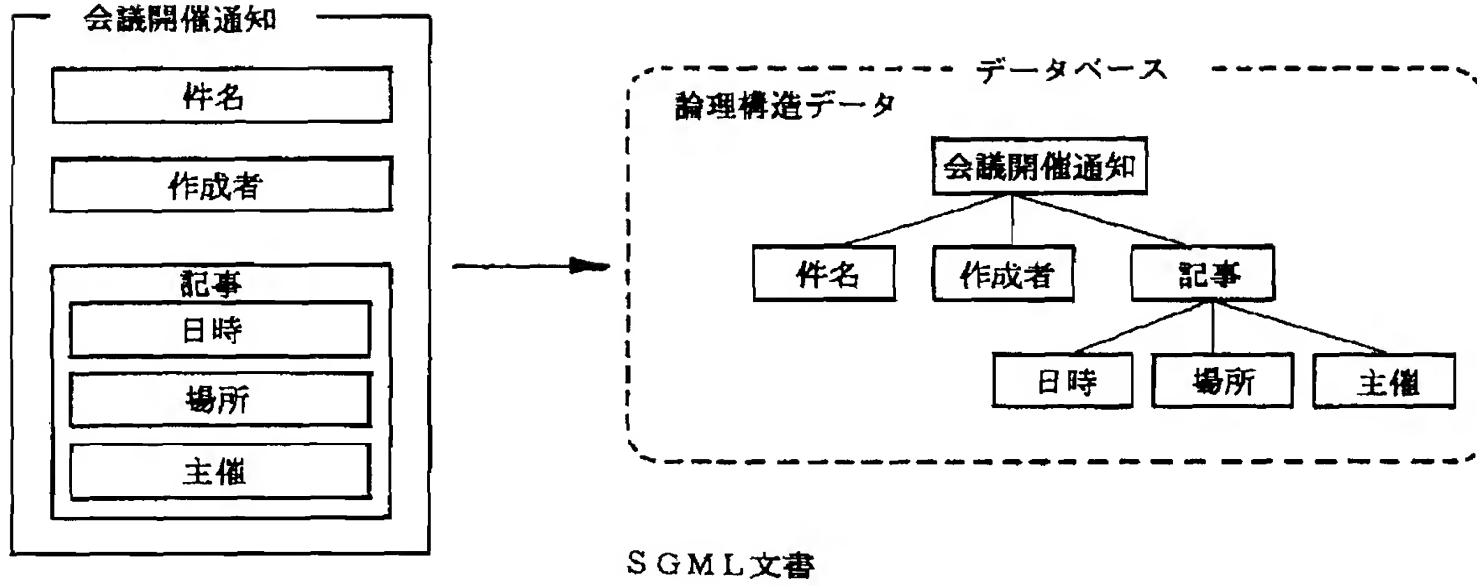
【図6】



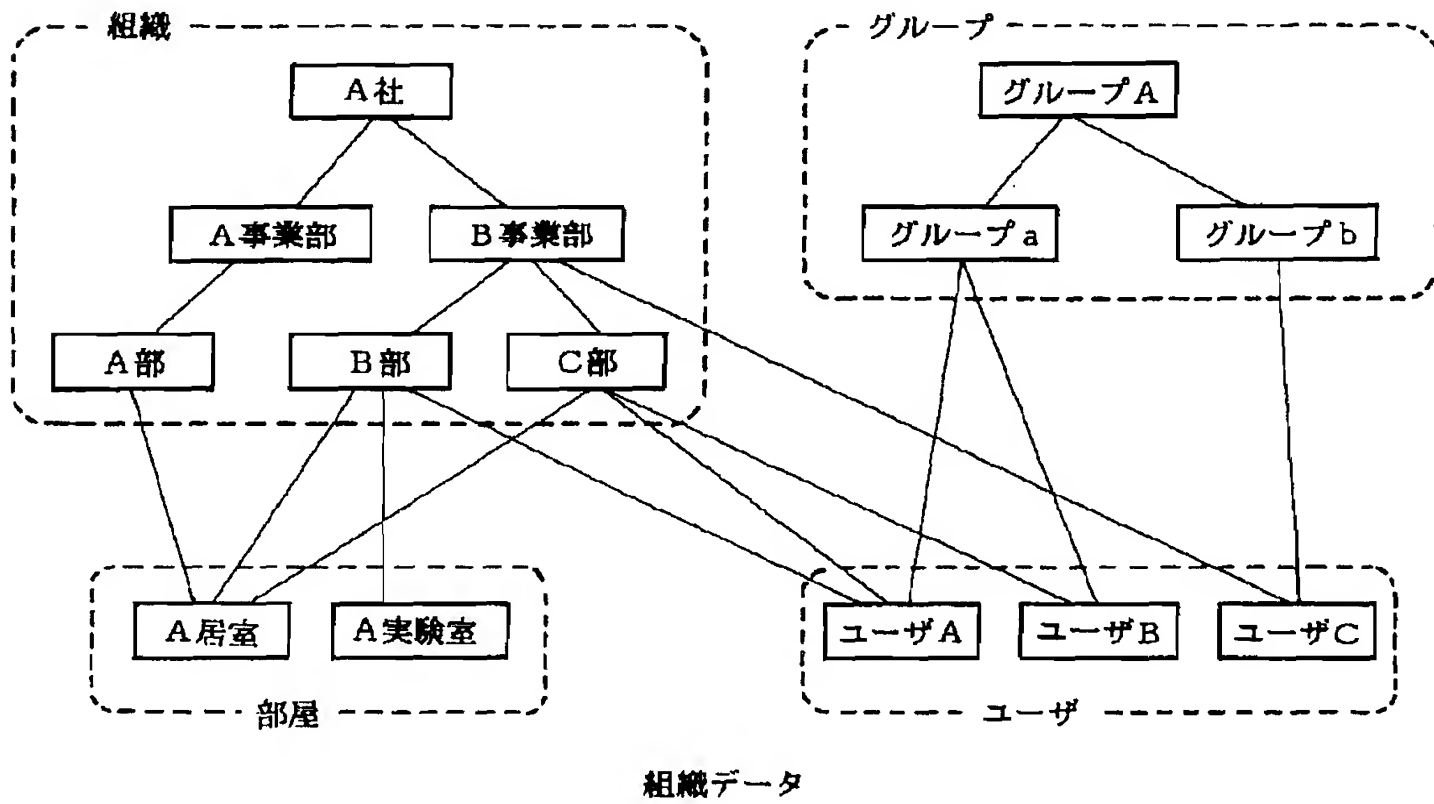


【図 2】

SGML 文書

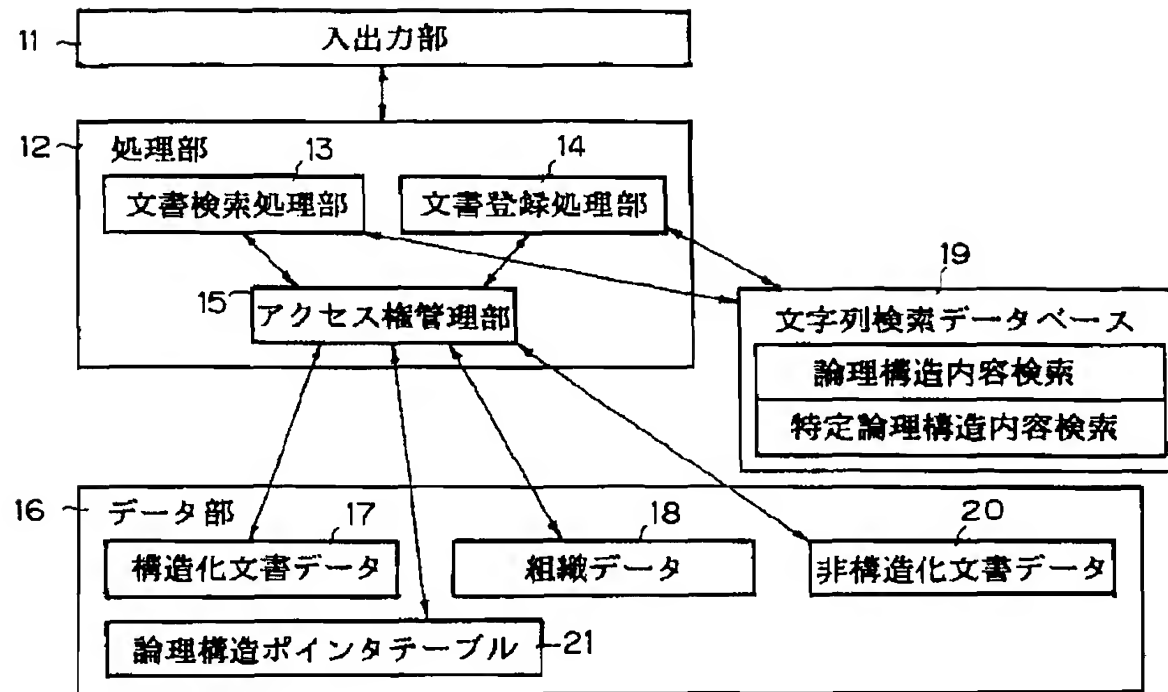


【図 3】

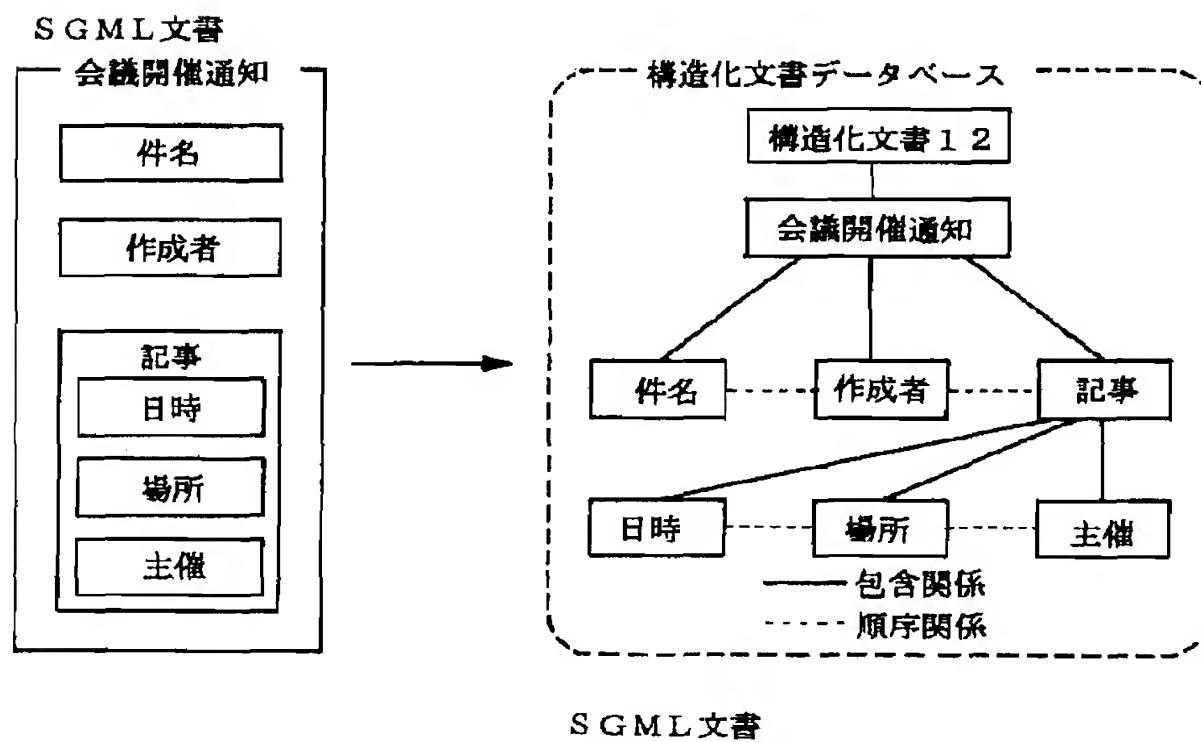


組織データ

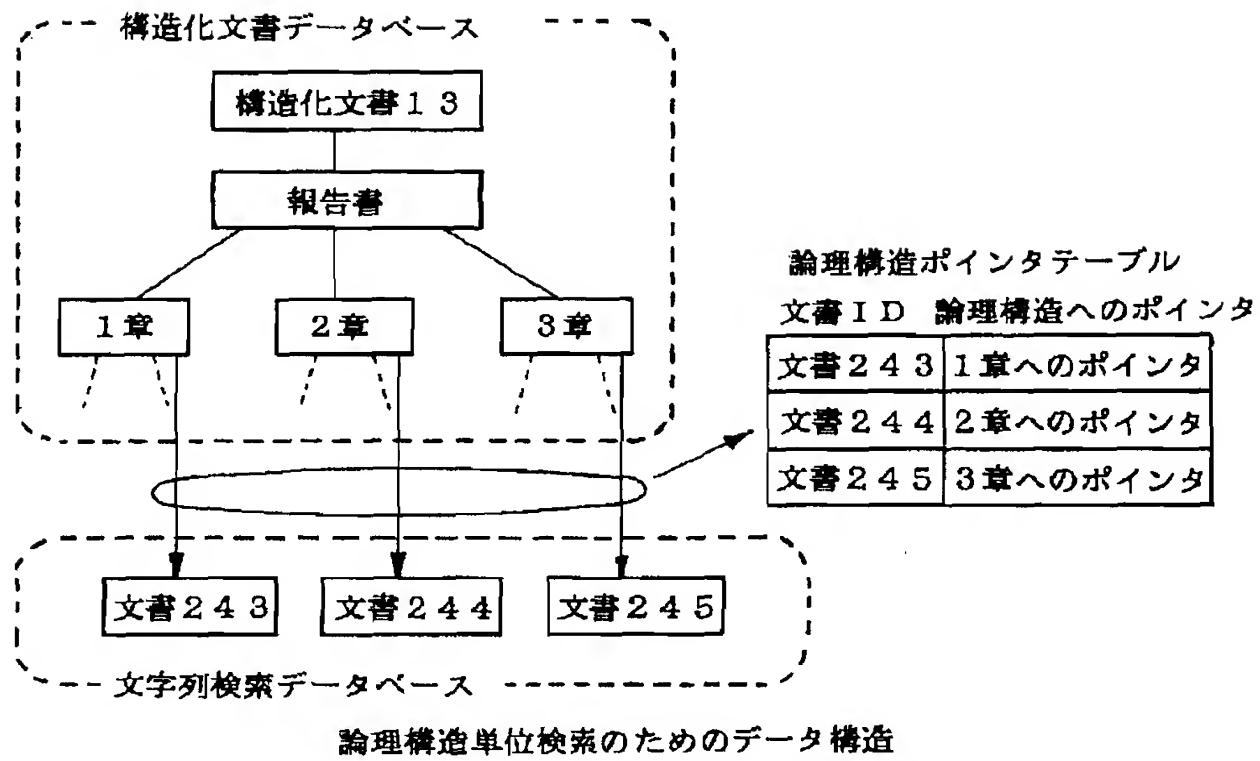
【図 4】



【図 5】



【図 7】



フロントページの続き

(51) Int. Cl. <sup>6</sup>

識別記号

庁内整理番号

F I

技術表示箇所

9194-5 L

G 0 6 F 15/40

3 2 0 B